

# Ensemble Kalman filtering with residual nudging: an extension to state estimation problems with nonlinear observation operators

XIAODONG LUO \*

*International Research Institute of Stavanger (IRIS), 5008 Bergen, Norway*

IBRAHIM HOTEIT

*King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

---

\* *Corresponding author address:* International Research Institute of Stavanger (IRIS), Thormøhlens Gate

55, 5008 Bergen, Norway

E-mail: xiaodong.luo@iris.no

## ABSTRACT

The ensemble Kalman filter (EnKF) is an efficient algorithm for many data assimilation problems. In certain circumstances, however, divergence of the EnKF might be spotted. In previous studies, the authors proposed an observation-space-based strategy, called residual nudging, to improve the stability of the EnKF when dealing with linear observation operators. The main idea behind residual nudging is to monitor and, if necessary, adjust the distances (misfits) between the real observations and the simulated ones of the state estimates, in the hope that by doing so one may be able to obtain better estimation accuracy.

In the present study, residual nudging is extended and modified in order to handle nonlinear observation operators. Such extension and modification result in an iterative filtering framework that, under suitable conditions, is able to achieve the objective of residual nudging for data assimilation problems with nonlinear observation operators. The 40 dimensional Lorenz 96 model is used to illustrate the performance of the iterative filter. Numerical results show that, while a normal EnKF may diverge with nonlinear observation operators, the proposed iterative filter remains stable and leads to reasonable estimation accuracy under various experimental settings.

## 1. Introduction

The ensemble Kalman filter (EnKF, see Burgers et al. 1998; Evensen 2006) and its variants (including, for example, Anderson 2001; Bishop et al. 2001; Hoteit et al. 2002; Luo and Moroz 2009; Pham 2001; Tippett et al. 2003; Wang et al. 2004; Whitaker and Hamill 2002) can be considered as Monte Carlo implementations of the celebrated Kalman filter

(Kalman 1960), in the sense that the mean and covariance of the Kalman filter are evaluated based on a finite (often small) number of samples of the underlying model states. Because of its ability to handle large-scale data assimilation problems, and its relative simplicity in implementation, the EnKF has received great attention from researchers in various fields.

In data assimilation, there are certain factors that may influence the performance of the EnKF. For instance, if the EnKF is implemented with a relatively small ensemble size, then the filter will often be subject to sampling errors. This may lead to some adverse effects (especially in high-dimensional models), including, for instance, under-estimation of the variances of state variables, over-estimation of the correlations between different state variables, and rank deficiency of the sample error covariance matrix (Whitaker and Hamill 2002; Hamill et al. 2009). In the literature, it is customary to adopt two auxiliary techniques, called covariance inflation (Anderson and Anderson 1999) and covariance localization (Hamill et al. 2001), to improve the performance of the EnKF. Intuitively, covariance inflation compensates for the under-estimated variances by artificially increasing it to some extent. It also increases the robustness of the EnKF from the point of view of  $H_\infty$  filtering theory (Luo and Hoteit 2011). Various methods of covariance inflation are proposed in the literature, for example, see Altaf et al. (2013); Anderson and Anderson (1999); Anderson (2007, 2009); Bocquet (2011); Bocquet and Sakov (2012); Luo and Hoteit (2011, 2013); Miyoshi (2011); Meng and Zhang (2007); Ott et al. (2004); Song et al. (2013); Triantafyllou et al. (2013); Whitaker and Hamill (2012); Zhang et al. (2004). On the other hand, covariance localization aims to taper the over-estimated correlations through, for instance, a Schur product between the sample error covariance matrix and a certain tapering matrix. In

effect, this also increases the rank of the sample error covariance matrix (Hamill et al. 2009).

Even equipped with both covariance inflation and localization, the EnKF may still suffer from filter divergence in certain circumstances, especially when there is substantial uncertainty, e.g., in terms of model and/or observation errors, in data assimilation problems (see, for example, the numerical results in Luo and Hoteit 2012). To mitigate filter divergence, in previous studies (Luo and Hoteit 2014, 2013, 2012) we considered a strategy, called data assimilation with residual nudging (DARN), which monitors and, if necessary, adjusts the distances (called residual norms) between the real observations and the simulated ones. Our numerical results showed that, under certain circumstances, a data assimilation algorithm equipped with residual nudging is not only more stable against filter divergence, but also performs better in terms of estimation accuracy.

The analytical and numerical results in Luo and Hoteit (2014, 2013, 2012) also show that, for linear observation operators, one is able to control the magnitudes of the residual norms under suitable conditions. An issue that we did not address yet is the nonlinearity in the observation operators. Our main motivation here is thus to fill this gap. To this end, we recast DARN as a least-squares problem and adopt an iterative filtering framework<sup>1</sup> to tackle the nonlinearity in the observation operators. Using this iterative filtering framework, one can achieve the objective of residual nudging under suitable conditions. For convenience, we refer to the observations from a linear (or nonlinear) observation operator as “linear observations” (or “nonlinear observations”), when it causes no confusion.

This work is organized as follows. Section 2 introduces the idea of DARN and outlines

---

<sup>1</sup>Here, by “iterative” we mean the presence of an iteration process (Eq. (11)) in each data assimilation cycle.

the method used in Luo and Hoteit (2013) for residual nudging with linear observations. In Section 3, the aforementioned method is extended and modified to tackle problems with nonlinear observations. In Section 4, various experiments are conducted to compare the proposed method with some existing algorithms in the literature. In addition, the stability of the proposed method is also investigated under different experimental settings. Finally, Section 5 concludes the work.

## 2. Residual nudging with linear observations

In this work we consider the state estimation problem in the following systems

$$\mathbf{x}_k = \mathcal{M}_{k,k-1}(\mathbf{x}_{k-1}) + \mathbf{u}_k, \quad (1a)$$

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \mathbf{v}_k, \quad (1b)$$

where  $\mathbf{x}_k$  is an  $m$ -dimensional model state and  $\mathbf{y}_k$  is the corresponding  $p$ -dimensional observation;  $\mathcal{M}_{k,k-1}$  is the model transition operator that maps the model state  $\mathbf{x}_{k-1}$  at time instant  $(k-1)$  to the next time instant  $k$ , and  $\mathcal{H}_k$  is the observation operator that projects the model state  $\mathbf{x}_k$  onto the observation space; and  $\mathbf{u}_k \in \mathbb{R}^m$  and  $\mathbf{v}_k \in \mathbb{R}^p$  are the model and observation errors, respectively. We assume that the observation error  $\mathbf{v}_k$  has zero mean and a non-singular covariance matrix  $\mathbf{R}_k$ . In the discussion below, the time index  $k$  is often uninvolved and thus dropped for ease of notation.

In this section, we focus on the case with linear observations. To this end, we rewrite the observation operator  $\mathcal{H}$  as  $\mathbf{H}$ . Suppose that  $\mathbf{y}^o$  is the real observation at a certain time instant, and  $\hat{\mathbf{x}}$  is the estimate of the underlying model state  $\mathbf{x}$ . Then the *residual*  $\hat{\mathbf{r}}$  is

defined as the difference between the simulated observation  $\mathbf{H} \hat{\mathbf{x}}$  and the real value  $\mathbf{y}^o$ , i.e.,  $\hat{\mathbf{r}} = \mathbf{H} \hat{\mathbf{x}} - \mathbf{y}^o$ .

Given a vector  $\mathbf{z} \in \mathbb{R}^p$  in the observation space, we use the weighted Euclidean norm

$$\|\mathbf{z}\|_{\mathbf{R}} \equiv \sqrt{\mathbf{z}^T \mathbf{R}^{-1} \mathbf{z}} \quad (2)$$

to measure its length. Under this setting, the objective of residual nudging is to find an analysis estimate  $\hat{\mathbf{x}}^a$ , such that the weighted Euclidean norm of the corresponding analysis residual  $\hat{\mathbf{r}}^a$  (the analysis residual norm hereafter) satisfies  $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \leq \beta_u \sqrt{p}$  for a pre-chosen scalar  $\beta_u$  ( $\beta_u > 0$ ). Readers are referred to Luo and Hoteit (2014, 2013, 2012) for the rationale behind this choice. To prevent over-fitting the observation, it may also be desirable to let  $\beta_l \sqrt{p} \leq \|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$  for another scalar  $\beta_l$  ( $0 < \beta_l < \beta_u$ ). Combining these constraints, the objective thus becomes

$$\beta_l \sqrt{p} \leq \|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \leq \beta_u \sqrt{p}. \quad (3)$$

Two methods were proposed in Luo and Hoteit (2014, 2013, 2012) for the purpose of residual nudging. In Luo and Hoteit (2014, 2012) it was suggested to solve a linear equation first, and then combine the resulting solution (called the “observation inversion”) with the original state estimate. In a follow-up work (Luo and Hoteit 2013), residual nudging was recast as a problem of choosing a proper covariance inflation factor, and some sufficient conditions in this regard were explicitly derived for the analysis residual norm to be bounded in the interval  $[\beta_l \sqrt{p}, \beta_u \sqrt{p}]$ . In the next section, the second method will be extended and modified to handle nonlinear observations. For this reason, in what follows, we summarize the method developed in Luo and Hoteit (2013). Readers are referred to Luo and Hoteit (2014, 2012) for more details about the first method.

In Luo and Hoteit (2013), a family of mean update formulae, in the form of

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G} (\mathbf{y}^o - \mathbf{H}\hat{\mathbf{x}}^b) , \quad (4a)$$

$$\mathbf{G} = \hat{\mathbf{C}}^b \mathbf{H}^T \left( \delta \mathbf{H} \hat{\mathbf{C}}^b \mathbf{H}^T + \gamma \mathbf{R} \right)^{-1} , \quad (4b)$$

is considered, where  $\hat{\mathbf{x}}^b$  is the sample mean of the background ensemble,  $\mathbf{G}$  is a gain matrix that may be considered as a slight generalization of the Kalman gain matrix in the EnKF. The coefficients  $\delta$  and  $\gamma$  in  $\mathbf{G}$  are some positive scalars whose ranges need to be determined in order for the analysis residual norm to be bounded in the desired interval (3), and  $\hat{\mathbf{C}}^b$  is a symmetric, positive semi-definite matrix. In general,  $\hat{\mathbf{C}}^b$  may be related, but not necessarily identical, to the sample error covariance matrix  $\hat{\mathbf{P}}^b$  of the background ensemble (to be further discussed below).

While the objective (3) can be achieved with the general gain matrix  $\mathbf{G}$  defined in Eq. (4b) (Luo and Hoteit 2013), it suffices to consider the conventional choice of  $\delta = 1$  here. In this case, the gain matrix  $\mathbf{G}$  resembles the Kalman gain in the EnKF, with  $1/\gamma$  being analogous to the multiplicative covariance inflation factor used in Anderson and Anderson (1999). With some algebra, it can be shown that the analysis residual norm satisfies (Luo and Hoteit 2013)

$$\|\hat{\mathbf{r}}^a\|_{\mathbf{R}} = \|\Phi (\mathbf{R}^{-1/2} \hat{\mathbf{r}}^b)\|_2 , \quad (5a)$$

$$\Phi = \gamma (\mathbf{A} + \gamma \mathbf{I}_p)^{-1} , \quad (5b)$$

$$\mathbf{A} = \mathbf{R}^{-1/2} \mathbf{H} \hat{\mathbf{C}}^b \mathbf{H}^T \mathbf{R}^{-T/2} , \quad (5c)$$

where  $\hat{\mathbf{r}}^b \equiv \mathbf{H}\hat{\mathbf{x}}^b - \mathbf{y}^o$  is the residual with respect to the background  $\hat{\mathbf{x}}^b$ ;  $\|\bullet\|_2$  denotes the standard Euclidean norm and satisfies  $\|(\mathbf{R}^{-1/2} \hat{\mathbf{r}}^b)\|_2 = \|\hat{\mathbf{r}}^b\|_{\mathbf{R}}$ , with  $\mathbf{R}^{1/2}$  being a square root

matrix of  $\mathbf{R}$ ; and  $\mathbf{I}_p$  represents the  $p$ -dimensional identity matrix. For ease of notation, in Eq. (5c) we have used  $\mathbf{R}^{-1/2}$  to denote the inverse of  $\mathbf{R}^{1/2}$ , and  $\mathbf{R}^{-T/2}$  to represent the transpose of  $\mathbf{R}^{-1/2}$ .

Given a matrix  $\mathbf{M}$  and a vector  $\mathbf{z}$  with suitable dimensions, one has the following inequalities (Grcar 2010)

$$\|\mathbf{M}\mathbf{z}\|_2 \leq \|\mathbf{M}\|_2 \|\mathbf{z}\|_2, \quad (6a)$$

$$\|\mathbf{M}^{-1}\|_2^{-1} \|\mathbf{z}\|_2 \leq \|\mathbf{M}\mathbf{z}\|_2. \quad (6b)$$

Applying (6) to the right-hand side of Eq. (5a) with  $\mathbf{M} = \Phi$  and  $\mathbf{z} = (\mathbf{R}^{-1/2}\hat{\mathbf{r}}^b)$ , one obtains the following sufficient conditions:

$$\frac{\xi_l}{1 - \xi_l} \lambda_{max} \leq \gamma \leq \frac{\xi_u}{1 - \xi_u} \lambda_{min}, \quad (7a)$$

$$\text{subject to } \beta_l \leq \frac{\beta_u}{\kappa + (1 - \kappa) \xi_u}, \quad (7b)$$

where  $\xi_l = \beta_l \sqrt{p} / \|\hat{\mathbf{r}}^b\|_{\mathbf{R}}$  and  $\xi_u = \beta_u \sqrt{p} / \|\hat{\mathbf{r}}^b\|_{\mathbf{R}}$ ,  $\lambda_{max}$  and  $\lambda_{min}$  are the maximum and minimum eigenvalues, respectively, of the matrix  $\mathbf{A}$  in Eq. (5c), and  $\kappa = \lambda_{max} / \lambda_{min}$  is the corresponding condition number. In case the observation size is large, such that it is expensive to evaluate the eigenvalues of  $\mathbf{A}$  directly, some alternative sufficient conditions can be obtained at a cheaper computational cost. This is omitted here for brevity and readers are referred to Luo and Hoteit (2013) for details. Finally, we note that, from Eq. (7), if only residual nudging is in consideration, then there is no restriction that  $\hat{\mathbf{C}}^b$  needs to be related to the sample error covariance  $\hat{\mathbf{P}}^b$ .

**Remark:** From the above deduction, one can relate residual nudging to certain forms of covariance inflation. As discussed in Luo and Hoteit (2011), a Kalman filter (or ensemble



Kalman filter) with covariance inflation is essentially a  $H_\infty$  filter (or its ensemble implementation, see Luo and Hoteit 2011). Compared with the Kalman filter (or its ensemble variants), the  $H_\infty$  filter (or its ensemble variants) puts more emphasis on the robustness of the estimation (Simon 2006). For more details of the similarities and differences between the Kalman and  $H_\infty$  filtering methods, readers are referred to Luo and Hoteit (2011) and the references therein.

### 3. Residual nudging with nonlinear observations

When the observation operator  $\mathcal{H}$  is nonlinear, residual nudging becomes more complicated, since an explicit relation between the analysis and background residuals, similar to that in Eq. (5) with linear observations, may no longer be available. On the other hand, if the state space of the dynamical model and the observation operator  $\mathcal{H}$  are continuous, and there is at least one model state  $\mathbf{x}^\dagger$  satisfying  $\mathcal{H}(\mathbf{x}^\dagger) = \mathbf{y}^o$  (or more generally, satisfying  $\|\mathcal{H}(\mathbf{x}^\dagger) - \mathbf{y}^o\|_{\mathbf{R}} \leq \beta_l \sqrt{p}$ ), and another model state whose residual norm is larger than  $\beta_u \sqrt{p}$ , then due to the continuity of  $\mathcal{H}$ , there exist model states satisfying (3), i.e., the objective of residual nudging is feasible.

In data assimilation practices, one often has an initial state estimate with a relatively large residual norm, while it may be more difficult to have readily available a state estimate with a sufficiently small residual norm. Therefore, in what follows, we present an iterative framework that aims to construct a sequence of model states with gradually decreasing residual norms as the iteration index increases. If the iteration process (see Eq. (11) later) is long enough, the residual norm may become sufficiently low such that (3) is satisfied.

*a. Iteration process to reduce the residual norm*

To establish the iteration process, we first note that in cases of linear observations, the analysis  $\hat{\mathbf{x}}^a$  in Eq. (4) (with  $\delta = 1$ ) is actually the solution of the following linear least-squares problem

$$\operatorname{argmin}_{\mathbf{x}} \|\mathbf{y}^o - \mathbf{H}\mathbf{x}\|_{\mathbf{R}}^2 + \gamma \|\mathbf{x} - \hat{\mathbf{x}}^b\|_{\hat{\mathbf{C}}^b}^2. \quad (8)$$

In line with this point of view, if the observation operator is nonlinear, then we aim to find a model state that solves the following nonlinear least-squares problem

$$\operatorname{argmin}_{\mathbf{x}} \|\mathbf{y}^o - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}}^2 + \gamma \|\mathbf{x} - \hat{\mathbf{x}}^b\|_{\hat{\mathbf{C}}^b}^2. \quad (9)$$

Some remarks regarding the cost function in (9) are in order. Firstly, for the objective (3) of residual nudging, it is intuitive to use only the first term (called the data mismatch term hereafter) in (9) as the cost function (see, for example, Kalnay and Yang 2010), which corresponds to the choice of  $\gamma = 0$  in (9). In many situations, minimizing the term  $\|\mathbf{y}^o - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}}^2$  alone corresponds to an ill-posed inverse problem, which has some notorious effects, e.g., non-uniqueness in the solutions and sensitivities of the solutions to (even tiny) observation errors (see, for example, the discussion in Engl et al. 2000). Therefore, in inverse problem theory and the geophysical data assimilation community, it is customary to introduce a certain regularization term, e.g.,  $\gamma \|\mathbf{x} - \hat{\mathbf{x}}^b\|_{\hat{\mathbf{C}}^b}^2$  in (9) ( $\gamma > 0$ ), with which the aforementioned problems can be avoided or mitigated. The presence of such a regularization term makes the solution of (9) approximately solve the minimization problem  $\operatorname{argmin}_{\mathbf{x}} \|\mathbf{y}^o - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}}^2$ , while the choice of  $\gamma$  follows a certain rule (e.g., (13) in this work).

In the literature, certain iteration processes are derived based on the cost function in (9) with  $\gamma = 1$ . As in the maximum likelihood ensemble filter (MLEF, see Zupanski 2005)

and other similar iterative ensemble filters (see, for example, Lorentzen and Nævdal 2011; Sakov et al. 2012), the rationale behind the choice of  $\gamma = 1$  may be largely explained from the point of view of Bayesian filtering, in the sense that the solution of (9) corresponds to the maximum a posteriori (MAP) estimate, when both the model state and the observation follow certain Gaussian distributions. However, from a practical point of view, such an interpretation may be only approximately valid in many situations. This is not only because the Gaussianity assumption may be invalid in many nonlinear dynamical models, but also because in reality it is often very challenging to accurately evaluate certain statistics, e.g., the error covariance matrices, of both the model state and the observation in large-scale problems.

With that said, in the iteration process below, we do not confine ourselves to a fixed cost function with either  $\gamma = 0$  or  $\gamma = 1$ . Instead, we let  $\gamma$  be adaptive with the iteration steps, which facilitates the gradual reduction of the residual norm of the state estimate, and is thus useful for the purpose of residual nudging. In Bocquet and Sakov (2012), an iteration process with essentially adaptive  $\gamma$  values is also introduced by combining the original inflation method in Bocquet (2011) and the iterative EnKF in Sakov et al. (2012). Note that in Bocquet and Sakov (2012) and Sakov et al. (2012) the cost functions are constructed with respect to the observations both at the present time (the so-called EnKF-N) and ahead in time (the so-called IEnKF-N) with respect to the model states to be optimized, while in the current work, the observations and the model states to be estimated are in the same assimilation cycles.

For convenience of discussion, let  $\{\hat{\mathbf{x}}^i\}$  ( $i = 0, 1, \dots$ ) be a sequence of state estimates obtained in the iteration process, with  $\hat{\mathbf{x}}^0$  equal to the background mean  $\hat{\mathbf{x}}^b$ , and  $\{\gamma^i\}$  ( $i =$

$0, 1, \dots$ ) a sequence of positive scalars that are associated with  $\{\hat{\mathbf{x}}^i\}$ . In the iteration process, we need to (a) calculate  $\hat{\mathbf{x}}^{i+1}$  based on  $\hat{\mathbf{x}}^i$  and  $\gamma^i$  at the previous iteration step, and (b) update the coefficient  $\gamma^i$  to a new value  $\gamma^{i+1}$ .

In this work, task (a) is undertaken by introducing a local linearization to the cost function in (9) at each iteration step, following Engl et al. (2000, ch. 11). More precisely, this involves linearizing the nonlinear operator  $\mathcal{H}$  locally around the most recent iterated estimate  $\hat{\mathbf{x}}^i$ , and replacing the reference point  $\hat{\mathbf{x}}^b$  in the regularization term  $\gamma\|\mathbf{x} - \hat{\mathbf{x}}^b\|_{\hat{\mathbf{C}}^b}^2$  with  $\hat{\mathbf{x}}^i$ . The latter modification is introduced in accordance with the former one, since in order to make local linearization approximately valid, it is expected that the new state estimate  $\hat{\mathbf{x}}^{i+1}$  should not be too far away from  $\hat{\mathbf{x}}^i$ . Using  $\hat{\mathbf{x}}^i$  as the reference point means that the observation  $\mathbf{y}^o$  is used multiple times in each data assimilation cycle. This choice may be justified by the fact that in many situations, the conventional non-iterative EnKF tends to be sub-optimal, due to, for instance, the nonlinearity in the dynamical model and/or the observation operator, the difficulties in accurately characterizing the statistics of the model and/or observation error(s), and the challenge in running the filter with a statistically sufficient ensemble size in large-scale applications. In such circumstances, assimilating the observations multiple times may help improve the filter's performance, as will be shown later (also see, for example, the numerical results in Luo and Hoteit 2012).

Taking the above considerations into account, at each iteration step, we solve a (local) minimization problem, in the form of

$$\underset{\mathbf{x}^{i+1}}{\operatorname{argmin}} \|\mathbf{y}^o - \mathcal{H}(\hat{\mathbf{x}}^i) - \mathbf{J}^i(\mathbf{x}^{i+1} - \hat{\mathbf{x}}^i)\|_{\mathbf{R}}^2 + \gamma^i \|\mathbf{x}^{i+1} - \hat{\mathbf{x}}^i\|_{\hat{\mathbf{C}}^b}^2, \quad (10)$$

where  $\mathcal{H}(\hat{\mathbf{x}}^i) + \mathbf{J}^i(\mathbf{x}^{i+1} - \hat{\mathbf{x}}^i)$  is the first order Taylor approximation of  $\mathcal{H}(\mathbf{x}^{i+1})$ , with  $\mathbf{J}^i$  being

the Jacobian matrix of  $\mathcal{H}$  at  $\hat{\mathbf{x}}^i$  (the evaluation of the Jacobian matrix is discussed later).

The solution  $\hat{\mathbf{x}}^{i+1}$  of the minimization problem is given by (Engl et al. 2000, ch. 11)

$$\hat{\mathbf{x}}^{i+1} = \hat{\mathbf{x}}^i + \mathbf{G}^i (\mathbf{y}^o - \mathcal{H}(\hat{\mathbf{x}}^i)) , \quad (11a)$$

$$\mathbf{G}^i = \hat{\mathbf{C}}^b(\mathbf{J}^i)^T \left( \mathbf{J}^i \hat{\mathbf{C}}^b(\mathbf{J}^i)^T + \gamma^i \mathbf{R} \right)^{-1} . \quad (11b)$$

Eq. (11) is similar to the iteration formula in Tarantola (2005, Eq. (3.49)) and the mean update formula in the ensemble square root filter (EnSRF, see Tippett et al. 2003), or more precisely, the ensemble transform Kalman filter (ETKF, see Bishop et al. 2001; Wang et al. 2004). In particular, if the observation operation  $\mathcal{H}$  is linear and  $\hat{\mathbf{C}}^b$  is equal to the background sample error covariance  $\hat{\mathbf{P}}^b$  (either with or without localization), then Eq. (11) is identical to the mean update formula in the ETKF with a multiplicative covariance inflation factor  $1/\gamma^i$ . In this sense, the mean update formula in the ETKF can be considered as a single step implementation of the iteration process in Eq. (11). In addition, one may further generalize Eq. (11) by introducing an additional scalar coefficient, say  $\alpha$ , in front of the gain matrix  $\mathbf{G}^i$ . Such an extension would then encompass the iteration processes of certain gradient-based optimization algorithms as special cases (see, for example, Eq. (A7) of Zupanski 2005, where  $\gamma^i \equiv 1$  during the iteration process).

Intuitively, with the first order Taylor approximation, one has

$$\begin{aligned} \|\mathcal{H}(\hat{\mathbf{x}}^{i+1}) - \mathbf{y}^o\|_{\mathbf{R}}^2 &\approx \|[\mathbf{I}_p - \mathbf{J}^i \hat{\mathbf{C}}^b(\mathbf{J}^i)^T (\mathbf{J}^i \hat{\mathbf{C}}^b(\mathbf{J}^i)^T + \gamma^i \mathbf{R})^{-1}] (\mathcal{H}(\hat{\mathbf{x}}^i) - \mathbf{y}^o)\|_{\mathbf{R}}^2 \\ &\leq \|(\mathcal{H}(\hat{\mathbf{x}}^i) - \mathbf{y}^o)\|_{\mathbf{R}}^2 , \end{aligned} \quad (12)$$

which implies that the residual norm of the estimate  $\hat{\mathbf{x}}^{i+1}$  at the  $(i+1)$ -th iteration tends to be no larger than that of  $\hat{\mathbf{x}}^i$  at the previous iteration. Similar to the situation with linear observations, in order for the inequality in (12) to be valid, there is no restriction that  $\hat{\mathbf{C}}^b$

be related to the background sample error covariance  $\hat{\mathbf{P}}^b$ .

From the point of view of the deterministic inverse problem theory, Eq. (11) can also be considered as an implementation of the regularized Levenberg-Marquardt method (see, for example, Engl et al. 2000, ch. 11), with the weight matrices for the data mismatch and regularization terms being  $\mathbf{R}$  and  $\hat{\mathbf{C}}^b$ , respectively. Compared with the conventional Levenberg-Marquardt method (which is already used in the data assimilation community, see Bocquet and Sakov 2012), there are a few differences in the regularized Levenberg-Marquardt method, which focuses more on the residual norm. For instance, in the regularized Levenberg-Marquardt method, it is customary to specify a parameter rule (see, for example, (13) below) to ensure the (local) convergence of the residual norm. In addition, the stopping criterion of the iteration process in Eq. (11) may normally involve a certain threshold of the residual norm, while in the conventional Levenberg-Marquardt method, the stopping criterion may instead be based on monitoring the relative change of the total cost function value (see, for example, Bocquet and Sakov 2012).

The relation between Eq. (11) and the deterministic inverse problem theory is useful for establishing the rule in choosing the parameter  $\gamma^i$ . To this end, we adopt the following parameter iteration rule (Engl et al. 2000, ch. 11)

$$\begin{aligned}\gamma^0 &> 0, \\ \gamma^{i+1} &= \rho^i \gamma^i, \text{ with } 1/r < \rho^i < 1 \text{ for some scalar } r > 1, \\ \lim_{i \rightarrow +\infty} \gamma^i &= 0,\end{aligned}\tag{13}$$

in which the scalar sequence  $\{\gamma^i\}$  gradually reduces to zero as  $i$  tends to  $+\infty$ , where the presence of the lower bound  $1/r$  for the coefficient  $\rho^i$  aims to prevent any abrupt drop-down

of  $\gamma^i$  to zero. When Eq. (11) is used in conjunction with Eq. (13), it can be analytically shown that the residual norms of the sequence of state estimates  $\{\hat{\mathbf{x}}^i\}$  converge to zero locally, provided that the equation  $\mathcal{H}(\mathbf{x}) = \mathbf{y}^o$  is solvable (and some other conditions are satisfied, see, e.g., Jin 2010). Of course, as discussed previously, it may not be desirable to have a too small residual norm in order to prevent over-fitting the observation. As a result, we choose to let the iteration process stop when either of the following two conditions are satisfied: (a) the residual norm of the state estimate is lower than a pre-set threshold  $\beta_u\sqrt{p}$  for the first time, in light of (3); or (b) the maximum iteration number is reached. Condition (b) is introduced here to control the runtime of the iteration process.

Figure 1 provides a schematic outline of the iteration process. Given a pair of quantities  $(\hat{\mathbf{x}}^{i-1}, \gamma^{i-1})$  at the  $(i-1)$ -th iteration step, Eqs. (11) and (13) are applied to update them to  $(\hat{\mathbf{x}}^i, \gamma^i)$ . Then, one checks if any of the above stopping conditions is met. If not, then Eqs. (11) and (13) are applied again to find new iterated values  $(\hat{\mathbf{x}}^{i+1}, \gamma^{i+1})$ , and so on, until at least one of the stopping conditions is satisfied. At the end of the iteration process, one obtains the final state estimate  $\hat{\mathbf{x}}^f$ , which will then be used to construct the analysis ensemble as in the ETKF (to be further explained in the next sub-section).

It is worth noting that Eq. (11) is similar to the iteration formulae used in Chen and Oliver (2013); Emerick and Reynolds (2012) in the context of the ensemble smoother (ES, see Evensen and van Leeuwen 2000), and in Stordal and Lorentzen (2014) in the context of the iterative adaptive Gaussian mixture (AGM) filter (Stordal et al. 2010). In Emerick and Reynolds (2012), a constraint,  $\sum_i 1/\gamma_i = 1$ , is imposed on the ES, implying that  $\gamma_i \geq 1$ . However, such a constraint may not guarantee that the data mismatch term in (9) can be sufficiently reduced. For instance, one can design a sequence  $\{\gamma^i\}$  with increasing values, e.g.,

$\gamma^{i+1} = \rho\gamma^i$  for some  $\rho > 1$  such that  $\{\gamma^i\}$  grows exponentially fast but still satisfies the constraint  $\sum_i 1/\gamma_i = 1$ . Then, if  $\{\gamma^i\}$  becomes large enough, the gain matrix  $\mathbf{G}^i$  in Eq. (11) tends to zero exponentially fast such that the iteration formula in Eq. (11a) would quickly make no significant change to the estimate  $\hat{\mathbf{x}}^i$  (results not shown). In Chen and Oliver (2013), the formula is established through an approximation, by discarding certain model state terms during the iteration process of the standard Levenberg-Marquardt algorithm. The parameters  $\gamma^i$  are determined through a way similar to the back-tracking line search method (Nocedal and Wright 2006) and may increase or decrease, depending on the circumstances. The convergence of the residual norms of the corresponding iteration process is, however, not clear yet. The iteration formula in Stordal and Lorentzen (2014) is similar to those in Chen and Oliver (2013); Emerick and Reynolds (2012), but is derived from the point of view of the Bayesian inversion theory. Under suitable conditions, asymptotic optimality can be achieved through the iteration formula in the sense of Stordal (2013).

*b. Implementation in the framework of the ETKF*

In this section, we consider incorporating the proposed iteration process (Eq.(11)) into the ETKF. The resulting filter is thus referred to as the iterative ETKF with residual nudging (IETKF-RN) hereafter. The idea here is to use the final model state  $\hat{\mathbf{x}}^f$  of the iteration process as the analysis, but introduce no modification to the analysis square root matrix of the normal ETKF. That is to say, given the same background ensemble, the normal ETKF and the IETKF-RN yield identical analysis square root matrices, but in general they may have different analyses. The analysis ensemble in the IETKF-RN is then produced based on



$\hat{\mathbf{x}}^f$  and the associated square root matrix, in exactly the same way as in the normal ETKF. The choice of the identical analysis square root matrix in both filters is motivated by the observation that, in the Kalman filter, the covariance update formula is independent of the prior or posterior mean (but not vice versa, see, for example, Simon 2006, ch. 5).

The remaining issues then involve specifying the following quantities in the iteration process: the covariance  $\hat{\mathbf{C}}^b$  and the Jacobian matrix  $\mathbf{J}^i$  in Eq. (11b), and the initial value  $\gamma^0$  and the reduction factor  $\rho^i$  in Eq. (13).

#### 1) SPECIFYING THE COVARIANCE $\hat{\mathbf{C}}^b$

To evaluate the gain matrix in Eq. (4b), one needs to compute the matrix product,  $\mathbf{J}^i \hat{\mathbf{C}}^b (\mathbf{J}^i)^T$ , which may be computationally expensive in large-scale problems. To alleviate this problem, we take advantage of the fact that, for the purpose of residual nudging, it is not mandatory to relate  $\hat{\mathbf{C}}^b$  to the background sample error covariance  $\hat{\mathbf{P}}^b$ . As a result, in our implementation,  $\hat{\mathbf{C}}^b$  is constructed based on a background matrix  $\mathbf{B}^{lt}$  that is the “climatological” covariance of a model trajectory from a long model run (see Section 4a on how  $\mathbf{B}^{lt}$  is obtained).  $\hat{\mathbf{C}}^b$  remains constant and thus avoids re-evaluation over time. To further reduce the computational cost, we let  $\hat{\mathbf{C}}^b$  be a diagonal matrix, whose diagonal elements correspond to those of  $\mathbf{B}^{lt}$ . By doing so, the diagonal elements of  $\hat{\mathbf{C}}^b$  may, on average, have magnitudes close to those of  $\hat{\mathbf{P}}^b$ , but require less time or storage when evaluating the relevant terms in Eq. (11b). On the other hand, if the computational cost is affordable, then it may be more preferable to use the full matrix  $\mathbf{B}^{lt}$ , or its hybrid with  $\hat{\mathbf{P}}^b$ , which may improve the estimation accuracy of the filter (results not shown).

**Remark:** If  $\hat{\mathbf{C}}^b$  is taken as a hybrid of  $\mathbf{B}^{lt}$  and  $\hat{\mathbf{P}}^b$ , then the IETKF-RN draws similarity to the hybrid EnKF-3DVar scheme (Hamill and Snyder 2000), in which the analysis mean at a given time instant is also obtained by minimizing a certain cost function, and the corresponding analysis ensemble is generated in the same way as in a certain EnKF (e.g., the EnKF with perturbed observations). As a further extension, one may also consider incorporating into (9) the observations in multiple assimilation cycles. In this case, the iterative framework is used as an (iterative) ensemble smoother (see, for example, Yang et al. 2012; Bocquet and Sakov 2013, 2014), and is similar to certain hybrid ensemble 4DVar schemes (see, for example, Liu et al. 2008).

## 2) EVALUATING THE JACOBIAN MATRIX $\mathbf{J}^i$

If the derivative of the observation operator  $\mathcal{H}$  is known, then  $\mathbf{J}^i$  can be explicitly constructed. In certain situations, although it is possible for one to evaluate the function values of  $\mathcal{H}$ , the function form itself may be too complex or even unknown to the user<sup>2</sup>. It is therefore challenging to obtain the analytic form of the Jacobian matrix. To this end, below we adopt a stochastic approximation method, called simultaneous perturbation stochastic approximation (SPSA, see, for example, Spall 1992), to approximate  $\mathbf{J}^i$ . The main reason for us to adopt the SPSA method is that it is a relatively simple approximation scheme. In real applications, however, one may replace the SPSA method by more accurate – but possibly also more sophisticated and expensive – approximation schemes.

In the SPSA method, to evaluate  $\mathbf{J}^i$  around  $\hat{\mathbf{x}}^i$ , a random perturbation  $\delta\mathbf{e} = (\delta e_1, \dots, \delta e_m)^T$

---

<sup>2</sup>Examples may include, for instance, neural networks or certain commercial softwares.

is first generated, where  $\delta e_j$  ( $j = 1, \dots, m$ ) takes the value 1 or  $-1$  with equal probability. Let  $\delta \mathbf{p} = \mathbf{S} \delta \mathbf{e}$ , where  $\mathbf{S}$  is an  $m \times m$  matrix, e.g., a square root of  $\hat{\mathbf{C}}^b$  as in our implementation. In addition, let the elements of  $\delta \mathbf{p}$  be  $\delta p_j$  ( $j = 1, \dots, m$ ), and define a corresponding point-wise inverse vector  $\delta \mathbf{p}_{inv} \equiv ((\delta p_1)^{-1}, \dots, (\delta p_m)^{-1})^T$ . Then, we calculate the approximate Jacobian from

$$\mathbf{J}^i \approx \frac{\mathcal{H}(\hat{\mathbf{x}}^i + \alpha \delta \mathbf{p}) - \mathcal{H}(\hat{\mathbf{x}}^i - \alpha \delta \mathbf{p})}{2\alpha} \times \delta \mathbf{p}_{inv}, \quad (14)$$

where  $\alpha$  is a scaling factor. Eq. (14) can be considered as a stochastic implementation of the finite difference scheme for Jacobian approximation. From this point of view,  $\alpha$  may take some relatively small value, e.g.,  $\alpha = 10^{-3}$  as in our implementation.

### 3) UPDATING THE PARAMETER $\gamma^i$

The initial value  $\gamma^0$  is chosen in a way such that relatively small changes are introduced to  $\hat{\mathbf{x}}^0$ , in light of the requirement of linearizing  $\mathcal{H}$  (through SPSA) in our implementation. To this end, in Eq. (11b), the terms  $\mathbf{J}^0 \hat{\mathbf{C}}^b (\mathbf{J}^0)^T$  and  $\gamma^0 \mathbf{R}$  are made comparable, by letting  $\gamma^0 = \text{trace}(\mathbf{J}^0 \hat{\mathbf{C}}^b (\mathbf{J}^0)^T) / \text{trace}(\mathbf{R})$ <sup>3</sup>, where  $\text{trace}(\bullet)$  denotes the trace of a matrix.

The deterministic inverse problem theory (see, for example, Engl et al. 2000, ch. 11) suggests that any parameter rule satisfying (13) is sufficient for the purpose of residual nudging. In our implementation, however,  $\mathbf{J}^i \hat{\mathbf{C}}^b (\mathbf{J}^i)^T$  may be singular. Therefore, if  $\gamma^i$  approaches zero too fast during the iteration (e.g., by letting  $\gamma^{i+1} = \rho \gamma^i$  for a constant  $\rho \in$

---

<sup>3</sup>If necessary, one may choose a larger value for  $\gamma^0$  (meaning a smaller step size in Eq. (11)), in order for a more accurate first order Taylor approximation in (10). A consequence of such a choice, however, is that more iteration steps may be needed to reduce the residual norm by the same amount.

$(0, 1)$ ), then one would quickly encounter numerical problems when inverting  $\mathbf{J}^i \hat{\mathbf{C}}^b (\mathbf{J}^i)^T + \gamma^i \mathbf{R}$  in Eq. (11b). To avoid this problem, in our implementation, we let  $\gamma^{i+1} = \gamma^i e^{-1/i}$  ( $i = 1, 2, \dots$ ). The reduction factor  $e^{-1/i}$  approaches 1 as the iteration index increases, while the parameters  $\gamma^i$  and  $\gamma^{i+1}$  still satisfy (13). In the same spirit, one may adopt similar parameter rules, e.g.,  $\gamma^{i+1} = \gamma^i (1 - 1/(i+1))$  ( $i = 1, 2, \dots$ ), which also worked well in our experiments (results not shown).

## 4. Experiments

### *a. Experimental settings*

The 40-dimensional L96 model (Lorenz and Emanuel 1998) is adopted as the testbed. The governing equations of the L96 model are given by

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, \dots, 40. \quad (15)$$

We define  $x_{-1} = x_{39}$ ,  $x_0 = x_{40}$  and  $x_{41} = x_1$  in Eq. (15).

The L96 model is integrated by the fourth-order Runge-Kutta method with a constant integration step of 0.05. In many of the experiments below, the following default settings are adopted unless otherwise stated: the L96 model is integrated from time 0 to 75 (Section 4b-1) or 525 (Section 4b-2) with the forcing term  $F = 8$ . To avoid the transition effect, the trajectory between 0 and 25 is discarded, and the rest (1000 and 10000 integration steps in Sections 4b-1 and 4b-2, respectively) is used as the truth in data assimilation. For convenience, we relabel the time step at 25.05 as step 1. The synthetic observation  $\mathbf{y}_k$  is obtained by measuring the odd number elements  $(x_{k,1}, x_{k,3}, \dots)$  of the state vector

$\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dots, x_{k,40}]^T$  every 4 time steps ( $k = 4, 8, 12, \dots$ ), in which the observation operator is given by  $\mathcal{H}(\mathbf{x}_k) = (f(x_{k,1}), f(x_{k,3}), \dots, f(x_{k,39}))^T$ , with  $f(x) = x^3/5$  being a cubic function. The observation error is assumed to follow a normal distribution,  $N(0, 1)$ , for each element in the observation vector. In some experiments, the forcing term  $F$ , the length of the assimilation time window, the frequency/density of the observations, the observation operator and so on, may be varied to investigate the sensitivities of the filter's performance to these factors.

To generate the initial background ensemble, we run the L96 model from 0 to 5000 (overall 100000 integration steps), and compute the temporal mean  $\mathbf{x}^{lt}$  and covariance  $\mathbf{B}^{lt}$  of the trajectory. We then assume that the initial state vector follows the normal distribution  $N(\mathbf{x}^{lt}, \mathbf{B}^{lt})$ , and draw a given number of samples as the initial background ensemble (which, of course, may not be the best possible way).  $\mathbf{B}^{lt}$  is also used to construct the matrix  $\hat{\mathbf{C}}^b$  in Eq. (11), with  $\hat{\mathbf{C}}^b = \text{diag}(\mathbf{B}^{lt})$  as previously mentioned, where  $\text{diag}(\mathbf{B}^{lt})$  stands for the diagonal matrix whose diagonal elements are those of  $\mathbf{B}^{lt}$ . In the experiments, the stopping conditions of the iteration process in Eq. (11) are either (a) when the residual norm is less than  $2\sqrt{p}$ , with  $p$  being the observation size; or (b) when the iteration number reaches the maximum of 15000 iterations. In some cases the maximum iteration number may also change.

In all the experiments below, neither covariance inflation, nor covariance localization, is applied to the IETKF-RN. The former choice is because, in the presence of parameter  $\gamma^i$  in Eq. (11), conducting extra covariance inflation is equivalent to changing the initial value  $\gamma^0$ , which is investigated in an experiment below. With regard to localization, our experience suggests that, in some cases (for example, that with the default experimental

settings at the beginning of this section and 20 ensemble members), conducting covariance localization may be beneficial for the IETKF-RN in the L96 model. In general, however, it is likely that the presence of covariance localization may alter the behaviour of IETKF-RN, in the sense that there is no guarantee any more that the iteration process (Eq. (11)), when equipped with covariance localization, moves along a residual-norm descent direction. Therefore, for our purpose, it appears more illustrative and conclusive for us to demonstrate only the performance of the IETKF-RN without localization.

### *b. Experiment results*

#### 1) A COMPARISON STUDY AMONG SOME ALGORITHMS

A comparison study is first conducted to investigate the performance of the IETKF-RN relative to the following algorithms: the normal ETKF (Bishop et al. 2001; Wang et al. 2004), the approximate Levenberg-Marquardt ensemble randomized maximum likelihood (LM-EnRML) method (Chen and Oliver 2013), and the iteration process of Eq. (11) with  $\gamma^i = 1 \forall i$  fixed during the iteration process (for distinction, we call this algorithm “IETKF-RN (constant  $\gamma$ )”). In the last algorithm, the iteration process aims to find a (local) minimum with respect to the cost function in Eq. (9) with  $\gamma = 1$ , which is essentially the same cost function adopted in, for example, the MLEF (Zupanski 2005). In this sense, the “IETKF-RN (constant  $\gamma$ )” algorithm can be considered as an alternative to the MLEF, with one of the differences from the MLEF being in the chosen optimization algorithm: in the MLEF, the conjugate gradient algorithm is adopted to minimize the cost function, while in the “IETKF-RN (constant  $\gamma$ )” algorithm, the Levenberg-Marquardt method is used instead. To

show the necessity of using adaptive  $\gamma$  values in certain circumstances, it would be desirable to conduct the comparison under the same conditions as far as possible. Therefore in what follows, we compare the IETKF-RN (with adaptive  $\gamma$ ) with the “IETKF-RN (constant  $\gamma$ )”, rather than directly with the MLEF.

It is also worth commenting on a difference between the iteration processes of the IETKF-RN and the LM-EnRML. In the LM-EnRML, the terms  $\hat{\mathbf{C}}^b(\mathbf{J}^i)^T$  and  $\mathbf{J}^i\hat{\mathbf{C}}^b(\mathbf{J}^i)^T$  in Eq. (11b) are replaced by  $\mathbf{S}_{x^i}(\mathbf{S}_{y^i})^T$  and  $\mathbf{S}_{y^i}(\mathbf{S}_{y^i})^T$ , respectively, where  $\mathbf{S}_{x^i}$  and  $\mathbf{S}_{y^i}$  are square root matrices of the sample covariances with respect to the ensemble  $\mathbf{X}^i \equiv \{\mathbf{x}_1^i, \dots, \mathbf{x}_n^i\}$  at the  $i$ -th iteration and the corresponding projection  $\mathbf{Y}^i \equiv \{\mathcal{H}(\mathbf{x}_1^i), \dots, \mathcal{H}(\mathbf{x}_n^i)\}$  onto the observation space, the same as those normally constructed in the EnKF (see, for example, Luo and Moroz 2009; Wang et al. 2004). With such substitutions, though, there is no guarantee that the iteration moves toward a direction along which the residual norm is reduced. This point will be illustrated later.

The normal ETKF is tested with the cubic observation function defined in the previous sub-section, with both covariance inflation and localization. In the experiments, we vary the inflation factor and half width of covariance localization within certain chosen ranges<sup>4</sup>, and we observe that the normal ETKF ends up with large root mean squared errors (RMSEs) in all tested cases, suggesting that the normal ETKF has in fact diverged. Divergences of the EnKF have also been reported in other studies with nonlinear observations, see, for example, Jardak et al. (2010).

Figure 2 reports the time series of residual norms (upper panel) and the corresponding

---

<sup>4</sup>Specifically the inflation factor  $\delta \in \{1.05, 1.1, 1.15, \dots, 1.30\}$ , and the half width  $l_c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ .

RMSEs (lower panel) obtained by applying the approximate LM-EnRM method with the same cubic observation function. The upper panel of Figure 2 plots the background residual norm (dash-dotted line) and that of the final iterative estimate (also called the final analysis hereafter) of the iteration process (solid line), together with the targeted upper bound (dashed line), which is the threshold  $\beta_u \sqrt{p} \approx 8.94$ , with  $\beta_u = 2$  and  $p = 20$  here. The time series of the final analysis residual norm overlaps with that of the background one in every assimilation cycle. They are thus indistinguishable in the figure. The reason for this result, in our opinion, is possibly that in this particular case, the approximate LM-EnRM method does not find an iterated estimate that is able to reduce the residual norm averaged over all ensemble members (a criterion used in Chen and Oliver 2013 in order to update the estimate). As a result, following the parameter rule in Chen and Oliver (2013), the  $\gamma$  value in the iteration formula continues to increase and eventually results in a negligible gain matrix, such that the final analysis estimate is essentially almost the same as the background. Consequently, in this case, there is almost no residual norm reduction. Instead, the background and final analysis residual norms appear identical and stay away from the targeted upper bound. For reference, the time series of the RMSEs of the final analysis estimates is also plotted in the lower panel of Figure 2, with the corresponding time mean RMSE being about 3.77.

Figure 3 plots the time series of the residual norms over the assimilation time window (upper panel); residual norm reduction of the iteration process at time step 500 (middle panel), an example that illustrates gradual residual norm reduction during the iteration process; and the time series of the corresponding RMSEs of the final estimates (lower panel), when the “IETKF-RN (constant  $\gamma$ )” algorithm is adopted to assimilate the cubic observations.



Compared with Figure 2, it is clear that in the upper panel of Figure 3, the residual norms of the final analysis estimates tend to be lower than the background ones in each assimilation cycle. In particular, in some cases, the final analysis residual norms approach, or even become slightly lower than, the pre-chosen upper bound of 8.94, while the corresponding initial background residual norms are often larger than 100. As a consequence of residual norm reduction, the corresponding time mean RMSE in the lower panel reduces to 3.38, smaller than that in Figure 2. Also note that the time series of the residual norms (upper panel) appears spiky. This may be because the estimation errors at certain time instants are relatively large (although the corresponding final analysis residual norms may have reasonable magnitudes). Consequently, after model propagation, the resulting background ensembles may have relatively large residual norms. In addition, the iteration process at those particular time instants may converge slowly, or may be trapped around certain local optima, such that the final analysis residual norms are only slightly lower than the background ones (hence the spikes). This phenomenon is also found in other experiments later.

Similar results (see Figure 4) are also observed when the iteration process in Eq. (11) is adopted, in conjunction with the adaptive parameter rule as described in Section 3b, to assimilate the cubic observations. One may see that the time mean RMSEs in Figs. 3 and 4 are close to each other. In this sense, it appears acceptable in this case to simply take  $\gamma^i = 1$  for all  $i$ , instead of adopting the more sophisticated parameter rule in Section 3b.

In what follows, though, we show with an additional example that the iteration process, when equipped with the adaptive parameter rule in Section 3b, tends to make the IETKF-RN more stable against filter divergence. To this end, we consider an exponential observation function  $f(x) = e^{x^2/10}$  that is applied to the odd number elements (i.e.,  $x_1, x_3, \dots, x_{39}$ )

of the state vector. For such strongly nonlinear observations, the “IETKF-RN (constant  $\gamma$ )” algorithm diverges after 30 time steps. In contrast, the IETKF-RN with adaptive  $\gamma^i$  appears to be more stable. As shown in the upper panel of Figure 5, it still manages to reduce the background residual norm in all assimilation cycles. In fact, at a few time steps, the final analysis residual norms also approach the targeted upper bound. Compared with the cubic observation case (see, for example, Figure 4), however, the final analysis residual norms appear much larger in many assimilation cycles, due to the stronger nonlinearity in the exponential observation operator.

The better stability of the IETKF-RN with adaptive  $\gamma$ , in comparison to the “IETKF-RN (constant  $\gamma$ )” algorithm in the case of exponential observations, may be understood from the optimization-theoretic point of view, when the iteration process in Eq.(11) is interpreted as a gradient-based optimization algorithm. For this type of optimization algorithm, it is usually suggested to start with a relatively small step size, so that the linearization involved in the algorithms may remain roughly valid (Nocedal and Wright 2006). In this regard, the IETKF-RN with adaptive  $\gamma$  may appear to be more flexible (e.g., one may make the initial step size small enough by choosing a large enough value for  $\gamma^0$ ), while there is no guarantee that the “IETKF-RN (constant  $\gamma$ )” algorithm may produce a small enough step size in general situations.

## 2) STABILITY OF THE IETKF-RN UNDER VARIOUS EXPERIMENTAL SETTINGS

Here we mainly focus on examining the stability of the IETKF-RN with adaptive  $\gamma$  under various experimental settings. To this end, in the experiments below, we adopt assimilation

time windows that are longer than those in the previous section. We note that the stability of the algorithm demonstrated below should be interpreted within the relevant experimental settings, and should not be taken for granted under different conditions (e.g., when with longer assimilation time windows).

Unless otherwise mentioned, in this section, the default experimental settings are as follows. The IETKF-RN is applied to assimilate cubic observations of the odd number state variables every 4 time steps, with the length of the assimilation time window being 10000 time steps. The variances of observation errors are 1. The IETKF-RN runs with 20 ensemble members and a maximum of 15000 iteration steps.

In the first experiment, we examine the performance of the IETKF-RN with both linear and nonlinear observations (linear observations are obtained by applying  $f(x) = x$  to specified state variables, plus certain observation errors). For either linear or nonlinear observations, there are two observation scenarios: one with all 40 state variables being observed (the full observation scenario), and the other with only the odd number state variables being observed (the half observation scenario). In each observation scenario, we consider the following four ensemble sizes: 5, 10, 15 and 20. For each ensemble size, we also vary the frequency, in terms of the number  $f_a$  of time steps, with which the observations are assimilated (i.e., the observations are assimilated every  $f_a$  time steps). In the experiment, the variances of observation errors are 1, and  $f_a$  is taken from the set  $\{1, 2, 4 : 4 : 60\}$ , where the notation  $v_i : \delta v : v_f$  is used to denote an array of numbers that grows from the initial value  $v_i$  to the final one  $v_f$ , with an even increment  $\delta v$  each time.

Figure 6 shows the time mean RMSEs (averaged over 10000 time steps) as functions of the ensemble size and the observation frequency, in the full and half observation scenarios,

respectively. In the full observation scenario (upper panels) and the half observation scenario with linear observations (panel (c)), for each ensemble size, the corresponding time mean RMSE appears to be a monotonically increasing function of the number  $f_a$  of time steps. On the other hand, when  $f_a$  is relatively small, it appears that the time mean RMSEs of all ensemble sizes are close to each other. As  $f_a$  increases, a larger ensemble size tends to yield a smaller time mean RMSE, although violations of this tendency may also be spotted in some cases of panel (b), possibly due to the sampling errors in the filter. In the half observation scenario with nonlinear observations (panel (d) of Figure 6), the behaviour of the IETKF-RN is similar to those in the other cases. There is, however, also a clear difference: instead of being a monotonically increasing function of  $f_a$ , the time mean RMSE in panel (d) exhibits V-shaped behaviour when  $f_a$  is relatively small, achieving the lowest value at  $f_a = 2$ , rather than at  $f_a = 1$  (possibly because the observations are over-fitted at  $f_a = 1$ ).

Overall, Figure 6 indicates that, the time mean RMSEs with linear observations (panels (a,c)) tend to be lower than those with nonlinear observations (panels (b,d)), suggesting that the nonlinearity in the observations may deteriorate the performance of the filter. On the other hand, the time mean RMSEs in the full observation scenarios (panels (a,b)) tend to be lower than those in the half observation scenarios (panels (c,d)). The latter may be explained from the point of view of solving a (linear or nonlinear) equation. In the full observation scenarios, at each time step that has an incoming observation  $\mathbf{y}^o$ , the number of state variables is equal to the observation size  $p$ . Therefore (provided that it is solvable), the equation  $\mathcal{H}(\mathbf{x}) = \mathbf{y}^o$  is well posed and has a unique solution. In this case, the smaller  $f_a$  is, the more observations (hence constraints) there are. Consequently, there are fewer degrees of freedom in constructing the solution of the equation, and this tends to drive the state

estimates toward the truth and yields relatively lower time mean RMSEs. In contrast, in the half observation scenario, the equation  $\mathcal{H}(\mathbf{x}) = \mathbf{y}^o$  is under-determined (ill-posed) and in general has non-unique solutions. In this case, a smaller  $f_a$  may tend to yield state estimates that better match the observations. However, similar to the issue of over-fitting observations in ill-posed problems, the smallest  $f_a$  does not necessarily result in the lowest possible time mean RMSE, as shown in panel (d).

As a side remark, we note that it is possible for one to further improve the performance of the IETKF-RN in Figure 6 with other experimental settings. For instance, for the half observation scenario with linear observations (panel (c)), if one lets  $\hat{\mathbf{C}}^b$  in Eq. (11b) be the sample covariance matrix of the background ensemble, and introduces both covariance inflation and localization to the filter, then in certain circumstances the time mean RMSE may be close to, or even lower than, 1 with relatively small  $f_a$  values<sup>5</sup>. However, for certain values of the half width of covariance localization, larger RMSEs or even filter divergence may also be spotted. We envision that this may be because in the presence of covariance localization, there is no guarantee that the iteration process (Eq. (11)) moves along a residual-norm descent direction. Therefore extra efforts are needed to take into account the impact of covariance localization on the search path of the iteration process, which will be investigated in the future.

We also follow Sakov et al. (2012) to test the IETKF-RN with a longer assimilation time window that consists of 100000 time steps. Here the half (nonlinear) observation scenario is investigated, with the ensemble size being 20 and the observation frequency being every

---

<sup>5</sup>For instance, when the inflation factor  $\delta = 0.08$ , the half width  $l_c = 0.1$  and  $\beta_u = 1$ , it is found that the time mean RMSEs of the IETKF-RN are around 0.50, 0.68 and 1.15, respectively, given  $f_a = 1, 2$  and 4.

4 time steps. Under these experimental settings, Figure 7 shows that the IETKF-RN runs stably, and its time mean RMSE is around 3.30, close to the values in the corresponding panels of Figs. 4 and 6.

Next, we test the performance of the IETKF-RN with different variances of observation errors. The experimental settings here are similar to those in panel (d) of Figure 6, except that the variances of observation errors are 0.01 and 10, respectively. As can be seen in Figure 8, for these two variances, the IETKF-RN also runs stably for all tested ensemble sizes and observation frequencies. Comparing panel (d) of Figure 6 and the panels of Figure 8, the IETKF-RN exhibits similar behaviours in these cases. It also indicates that when  $f_a$  is relatively small (say,  $f_a = 4$ ), smaller variances tend to lead to lower time mean RMSEs; while when  $f_a$  is relatively large (say,  $f_a = 60$ ), the situation seems to be the opposite. Since both  $f_a$  and the variances of observation errors affect the quality of the subsequent background ensembles, we conjecture that the above phenomenon occurs because for different combinations of  $f_a$  and variances, different *relative weights* are assigned to the background ensembles and the observations at the analysis steps. As a result, when  $f_a$  is relatively large, smaller variances do not necessarily lead to lower time mean RMSEs. Similar results are also found in, for example, Luo and Hoteit (2014, Figure 7).

We also investigate the effect of the maximum number of iterations on the performance of the IETKF-RN. The set of maximum numbers of iterations tested in the experiment is  $\{1, 10, 100, 1000, 10000, 100000\}$ . Figure 9 shows the time mean RMSE of the IETKF-RN as a function of the maximum number of iterations. There, one can see that the time mean RMSE tends to decrease as the maximum number of iterations increases.

Finally, we examine the impacts of (potentially) mis-specifying the forcing term  $F$  in the

L96 model and/or the variances of observation errors, on the performance of the IETKF-RN. In the experiment, the true forcing term  $F$  is 8, and the true variances of observation errors are 1 for all elements of the observations. The tested  $F$  values are taken from the set  $\{4 : 2 : 12\}$ , and the tested variances of observation errors are  $\{0.25, 0.5, 1, 2, 5, 10\}$ . Figure 10 reports the time mean RMSE as functions of the forcing term  $F$  and the variances of observation errors. One can see that the time mean RMSE seems not very sensitive to the (potential) mis-specification of the variances of observation errors, possibly because with cubic observations,  $\mathbf{J}^i \hat{\mathbf{C}}^b (\mathbf{J}^i)^T$  in Eq. (11) dominates  $\gamma^i \mathbf{R}$  for a moderate initial  $\gamma$  value. Therefore a mis-specification of  $\mathbf{R}$  in the experiments may not have a substantial impact on the iteration process. If one increases the initial  $\gamma$  value, or assimilates linear observations instead, then there can be more variations in the final estimation errors (results not shown).

The performance of the IETKF-RN in Figure 10, on the other hand, does appear to be sensitive to the potential mis-specification of  $F$ . Interestingly, for all tested variances of observation errors, the filter’s best performance is obtained at  $F = 6$ , rather than  $F = 8$ <sup>6</sup>. This suggests that, in certain situations, the filter might actually achieve better performance in the presence of certain suitable model errors, rather than with the perfect model. Similar observations are also reported in the literature, e.g., Gordon et al. (1993); Whitaker and Hamill (2012), in which it is found that introducing certain artificial model errors may in fact improve the performance of a data assimilation algorithm. Overall, Figure 10 suggests that the IETKF-RN can also run stably even with substantial uncertainty in the system.

---

<sup>6</sup>In the full observation scenario, however, the lowest time mean RMSE is indeed achieved at  $F = 8$  (results not shown).

## 5. Conclusion

In this work, we introduced the concept of data assimilation with residual nudging. Based on the method derived in a previous study, we proposed an iterative filtering framework to handle nonlinear observations in the context of residual nudging. The proposed iteration process is related to the regularized Levenberg-Marquardt algorithm from inverse problem theory. Such an interpretation motivated us to implement the proposed algorithm with an adaptive coefficient  $\gamma$ .

For demonstration, we implemented an iterative filter based on the ensemble transform Kalman filter (ETKF). Numerical results showed that the resulting iterative filter exhibited remarkable stability in handling nonlinear observations under various experimental settings, and that the filter achieved reasonable performance in terms of root mean squared errors.

For data assimilation in large-scale problems, it may not be realistic to conduct a large number of iterations, because of the limitation in computational resources. In this regard, one topic in our future research is to explore the possibility of enhancing the convergence rate of the iterative filter.

## Acknowledgment

We would like to thank three anonymous reviewers for their constructive comments and suggestions. The first author would also like to thank the IRIS/CIPR cooperative research project “Integrated Workflow and Realistic Geology” which is funded by industry partners ConocoPhillips, Eni, Petrobras, Statoil, and Total, as well as the Research Council of Norway



(PETROMAKS) for financial support.

## REFERENCES

- Altaf, U. M., T. Butler, X. Luo, C. Dawson, T. Mayo, and H. Hoteit, 2013: Improving short range ensemble Kalman storm surge forecasting using robust adaptive inflation. *Mon. Wea. Rev.*, **141**, 2705–2720, doi:10.1175/MWR-D-12-00310.1.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Anderson, J. L., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59A** (2), 210–224.
- Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83.
- Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

- Bocquet, M., 2011: Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, **18** (5), 735–750, doi:10.5194/npg-18-735-2011.
- Bocquet, M. and P. Sakov, 2012: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes in Geophysics*, **19** (3), 383–399, doi:10.5194/npg-19-383-2012.
- Bocquet, M. and P. Sakov, 2013: Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlin. Processes Geophys.*, **20** (5), 803–818.
- Bocquet, M. and P. Sakov, 2014: An iterative ensemble Kalman smoother. *Quart. J. Roy. Meteor. Soc.*, in press.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: On the analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Chen, Y. and D. Oliver, 2013: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, **17**, 689–703.
- Emerick, A. A. and A. C. Reynolds, 2012: Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, **55**, 3–15.
- Engl, H. W., M. Hanke, and A. Neubauer, 2000: *Regularization of Inverse Problems*. Springer.
- Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. Springer.

- Evensen, G. and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings F in Radar and Signal Processing*, **140**, 107–113.
- Grcar, J. F., 2010: A matrix lower bound. *Linear Algebra and its Applications*, **433**, 203–220.
- Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- Hamill, T. M., J. S. Whitaker, J. L. Anderson, and C. Snyder, 2009: Comments on “Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems”. *J. Atmos. Sci.*, **66**, 3498–3500.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.
- Hoteit, I., D. T. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *Journal of Marine Systems*, **36**, 101–127.
- Jardak, M., I. M. Navon, and M. Zupanski, 2010: Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation. *Int. J. Num. Methods Fluids*, **62**, 374–402.

- Jin, Q., 2010: On a regularized Levenberg–Marquardt method for solving nonlinear inverse problems. *Numerische Mathematik*, **115** (2), 229–259.
- Kalman, R., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.*, **82**, 35–45.
- Kalnay, E. and S.-C. Yang, 2010: Accelerating the spin-up of ensemble kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **136** (651), 1644–1651.
- Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136** (9), 3363–3373.
- Lorentzen, R. and G. Nævdal, 2011: An iterative ensemble Kalman filter. *IEEE Transactions on Automatic Control*, **56**, 1990–1995.
- Lorenz, E. N. and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.
- Luo, X. and I. Hoteit, 2011: Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 3938–3953.
- Luo, X. and I. Hoteit, 2012: Ensemble Kalman filtering with residual nudging. *Tellus A*, **64**, 17130, open access, doi:10.3402/tellusa.v64i0.17130.
- Luo, X. and I. Hoteit, 2013: Covariance inflation in the ensemble Kalman filter: a residual nudging perspective and some implications. *Mon. Wea. Rev.*, **141**, 3360–3368, doi:10.1175/MWR-D-13-00067.1.

- Luo, X. and I. Hoteit, 2014: Efficient particle filtering through residual nudging. *Quart. J. Roy. Meteor. Soc.*, **140**, 557–572, doi:10.1002/qj.2152.
- Luo, X. and I. M. Moroz, 2009: Ensemble Kalman filter with the unscented transform. *Physica D*, **238**, 549–562.
- Meng, Z. and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135** (4), 1403–1423.
- Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535.
- Nocedal, J. and S. J. Wright, 2006: *Numerical optimization*. 2d ed., Springer.
- Ott, E., et al., 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.
- Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative EnKF for strongly nonlinear systems. *Mon. Wea. Rev.*, **140** (6), 1988–2004.
- Simon, D., 2006: *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Wiley-Interscience, 552 pp.
- Song, H., I. Hoteit, B. D. Cornuelle, X. Luo, and A. C. Subramanian, 2013: An adjoint-based adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **141** (2013), 3343–3359.

- Spall, J. C., 1992: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *Automatic Control, IEEE Transactions on*, **37** (**3**), 332–341.
- Stordal, A. S., 2013: A new approximate algorithm for Bayesian sampling in high dimension. submitted.
- Stordal, A. S., H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. Vallès, 2010: Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, **15**, 293–305.
- Stordal, A. S. and R. J. Lorentzen, 2014: An iterative version of the adaptive Gaussian mixture filter. *Computational Geosciences*, in press, doi:10.1007/s10596-014-9402-6.
- Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Triantafyllou, G., I. Hoteit, X. Luo, K. Tsias, and G. Petihakis, 2013: Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan sea. *Journal of Marine Systems*, **125**, 90 – 100, doi:10.1016/j.jmarsys.2012.12.006.
- Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered simplex ensemble. *Mon. Wea. Rev.*, **132**, 1590–1605.
- Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

- Whitaker, J. S. and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089.
- Yang, S.-C., E. Kalnay, and B. Hunt, 2012: Handling nonlinearity in an ensemble Kalman filter: Experiments with the three-variable Lorenz model. *Mon. Wea. Rev.*, **140**, 2628–2646.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132** (5), 1238–1253.
- Zupanski, M., 2005: Maximum likelihood ensemble filter: theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.

# List of Figures

- 1     A schematic description of the iteration process. 42
  
- 2     Time series of the residual norm and the corresponding RMSE of the approximate LM-EnRML method for the cubic observation operator. Note that in the upper panel, the time series of the background residual norm (dash-dotted) overlaps with that of the final analysis one (solid). 43
  
- 3     IETKF-RN with fixed  $\gamma = 1$  applied to cubic observations. (a) Upper panel: time series of residual norms over the assimilation time window; (b) Middle panel: residual norm (solid) reduction of the iteration process at time step 500, an example used to illustrate how the residual norms of the iterative estimates are gradually reduced toward the targeted upper bound (dash-dotted); (c) Lower panel: time series of RMSEs of the final analysis estimates over the assimilation time window. 44
  
- 4     As in Figure 3, except that now the IETKF-RN is associated with adaptive  $\gamma$  during the iteration process. 45
  
- 5     As in Figure 3, but now the IETKF-RN with adaptive  $\gamma$  is applied to exponential observations. 46



- 6 Time mean RMSEs of the IETKF-RN as functions of the observation frequency (in number of time steps), for different ensemble sizes. The upper panels (a,b) are for the full observation scenario ( $p=40$ ), and the lower ones (c,d) are for the half observation scenario ( $p=20$ ). On the other hand, the left panels (a,c) are with linear observations, while the right ones (b,d) are with nonlinear observations. In all the above cases, the variances of observation errors are 1. Note that for the cases with linear observations,  $\beta_u$  is set to 1. In addition, the corresponding Jacobians can be calculated exactly, so there is no need to apply SPSA to approximate them. 47
- 7 As in Figure 4, except that the length of the assimilation time window is now 100000 time steps. For visualization, in the upper and lower panels we show only the time series between the time steps 50000 and 51000. The average time mean RMSE in the legend of the lower panel is, however, calculated with respect to the whole assimilation time window. 48
- 8 Time mean RMSEs of the IETKF-RN as functions of the observation frequency (in number of time steps), for different ensemble sizes. Variances of observation errors in the upper panel are 0.01, and those in the lower panel are 10. 49
- 9 Time mean RMSEs of the IETKF-RN as a function of the maximum number of iterations. With different orders of the magnitudes of the maximum iteration numbers, the logarithmic scale is used for the horizontal axis. 50

- 10 Time mean RMSE of the IETKF-RN as functions of the potentially mis-  
specified forcing term  $F$  and the variances of observation errors. For visual-  
ization, here the logarithmic scale is also used for the horizontal axis. 51

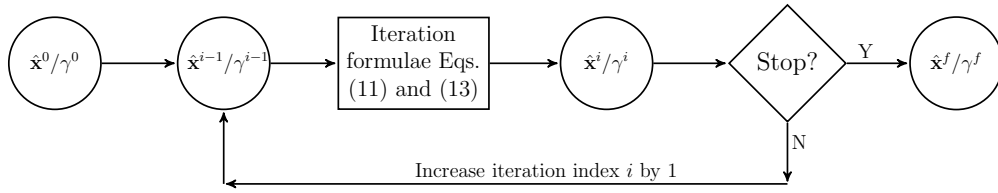
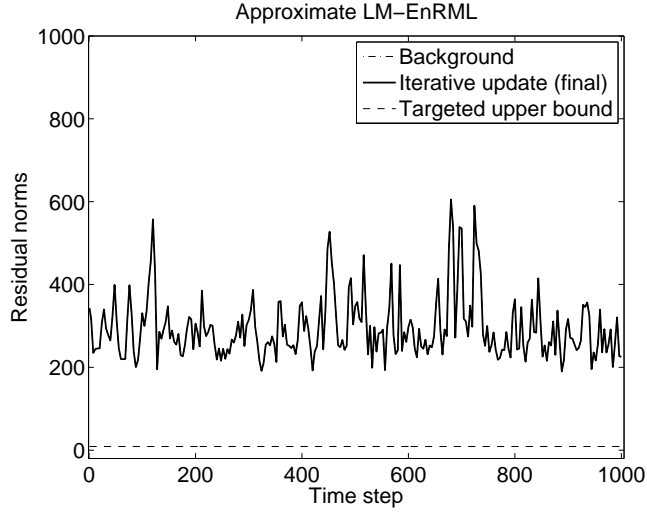
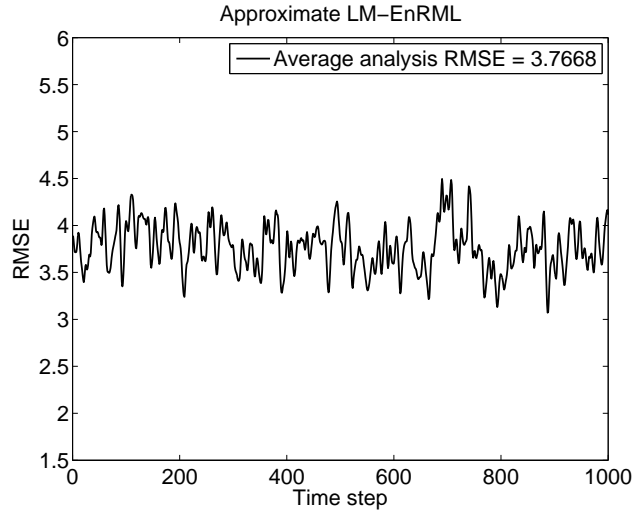


FIG. 1. A schematic description of the iteration process.

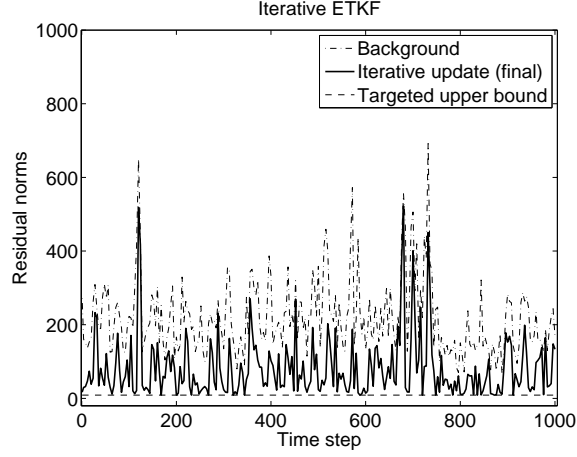


(a) Time series of residual norm

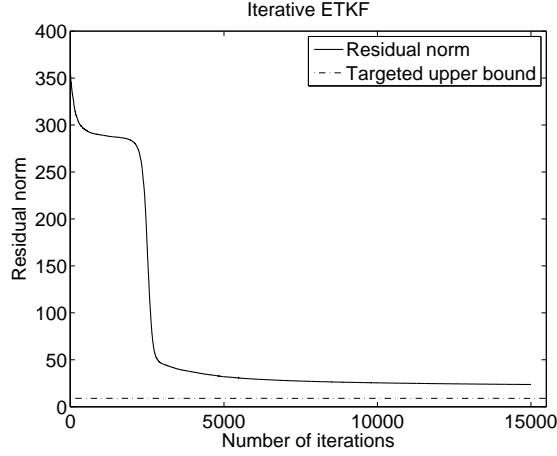


(b) Time series of RMSE

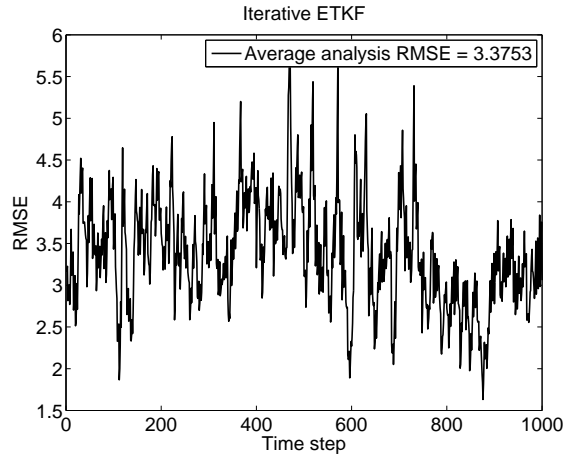
FIG. 2. Time series of the residual norm and the corresponding RMSE of the approximate LM-EnRML method for the cubic observation operator. Note that in the upper panel, the time series of the background residual norm (dash-dotted) overlaps with that of the final analysis one (solid).



(a) Time series of residual norm

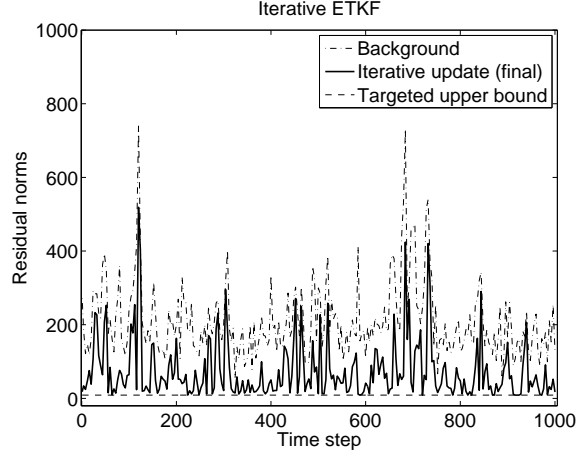


(b) Residual norm reduction of the iteration process at time step 500

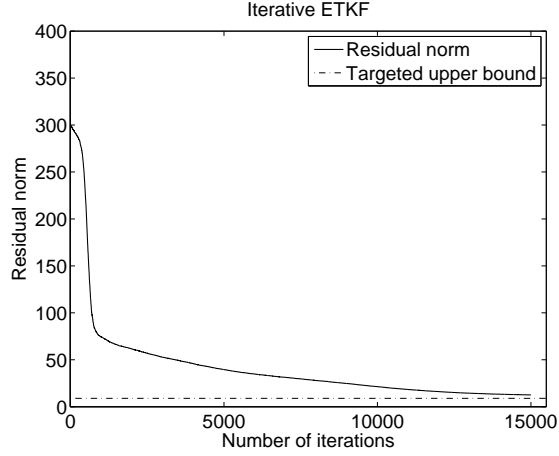


(c) Time series of RMSE

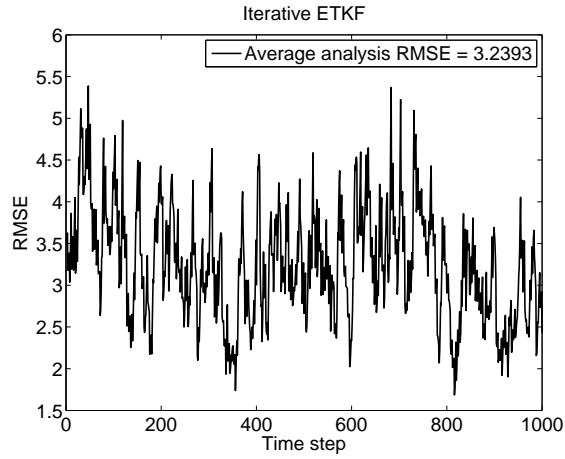
FIG. 3. IETKF-RN with fixed  $\gamma = 1$  applied to cubic observations. (a) Upper panel: time series of residual norms over the assimilation time window; (b) Middle panel: residual norm (solid) reduction of the iteration process at time step 500, an example used to illustrate how the residual norms of the iterative estimates gradually reduced toward the targeted upper bound (dash-dotted); (c) Lower panel: time series of RMSEs of the final analysis estimates over the assimilation time window.



(a) Time series of residual norm

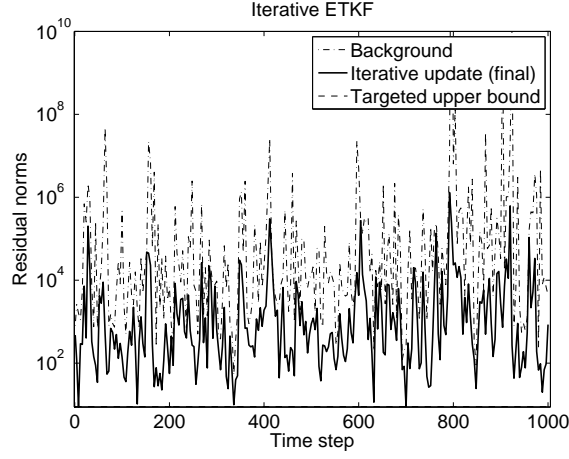


(b) Residual norm reduction of the iteration process at time step 500

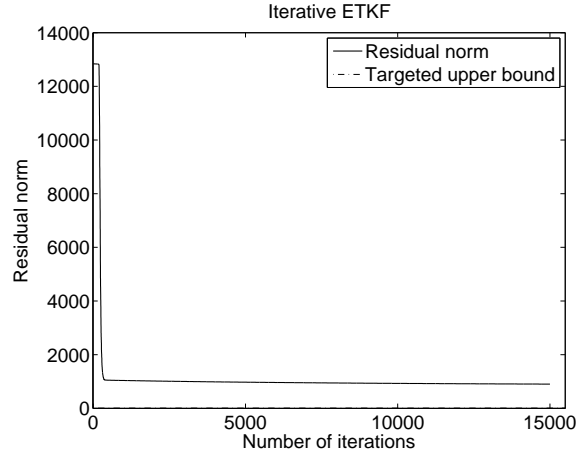


(c) Time series of RMSE

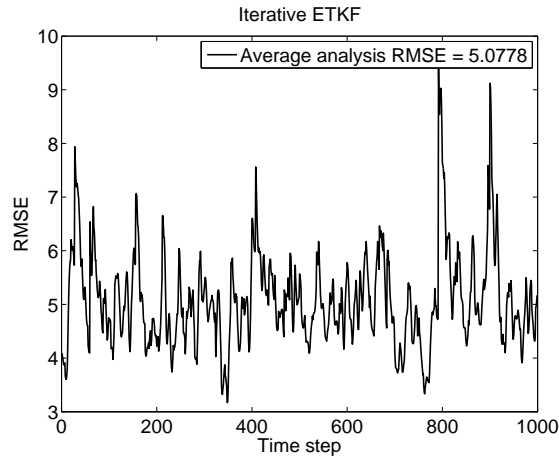
FIG. 4. As in Figure 3, except that now the IETKF-RN is associated with adaptive  $\gamma$  during the iteration process.



(a) Time series of residual norm

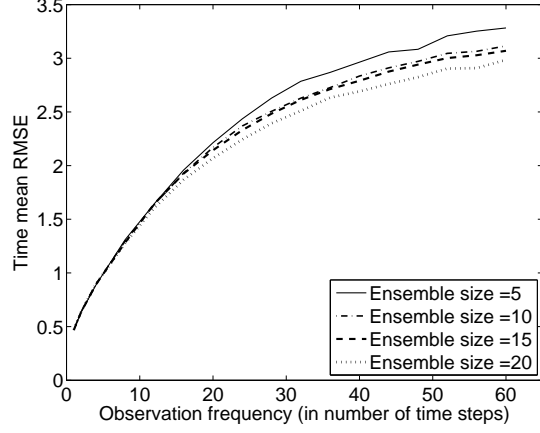


(b) Residual norm reduction of the iteration process at time step 500

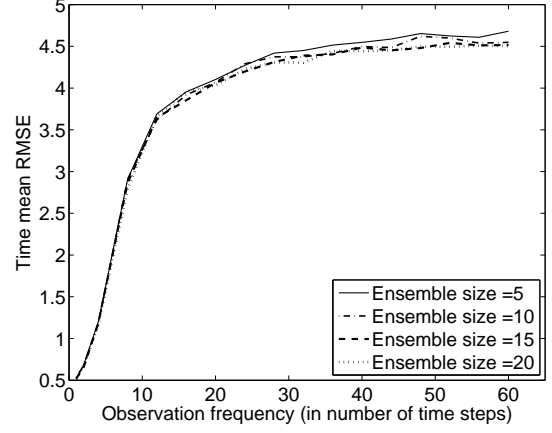


(c) Time series of RMSE

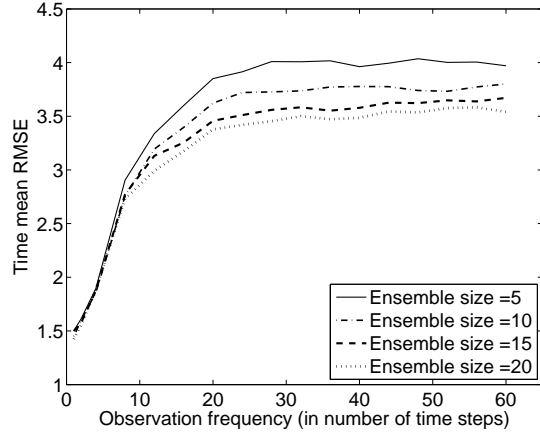
FIG. 5. As in Figure 3, but now the IETKF-RN with adaptive  $\gamma$  is applied to exponential observations.



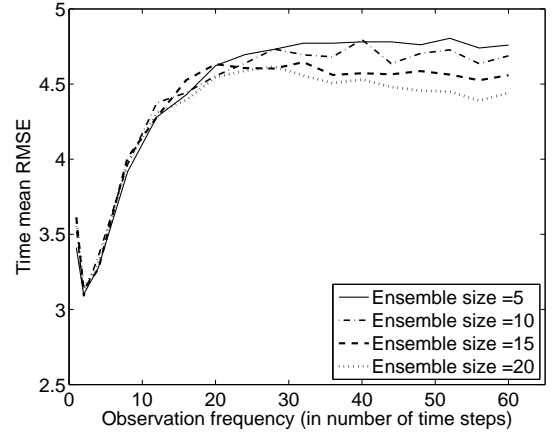
(a) Full (linear) observation scenario



(b) Full (nonlinear) observation scenario



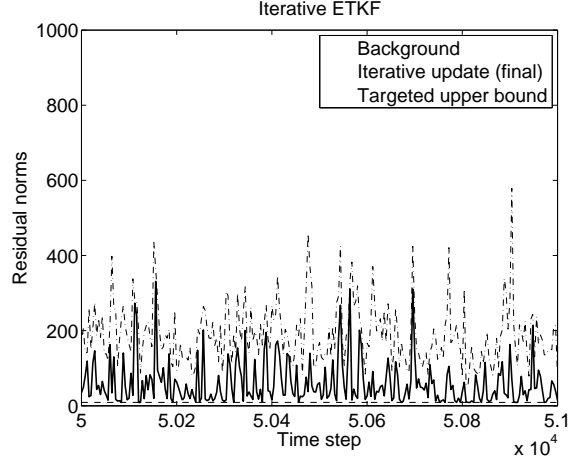
(c) Half (linear) observation scenario



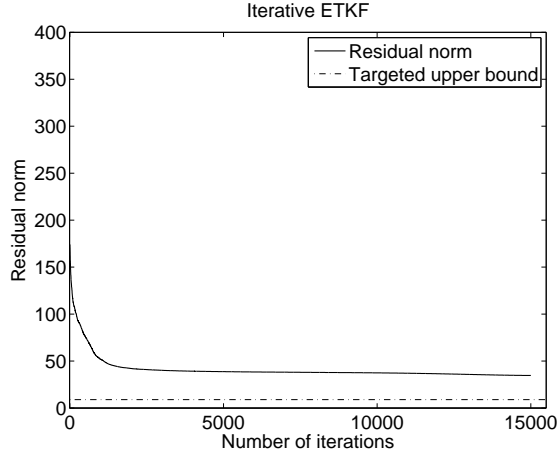
(d) Half (nonlinear) observation scenario

FIG. 6. Time mean RMSEs of the IETKF-RN as functions of the observation frequency (in number of time steps), for different ensemble sizes. The upper panels (a,b) are for the full observation scenario ( $p=40$ ), and the lower ones (c,d) are for the half observation scenario ( $p=20$ ). On the other hand, the left panels (a,c) are with linear observations, while the right ones (b,d) are with nonlinear observations. In all the above cases, the variances of observation errors are 1. Note that for the cases with linear observations,  $\beta_u$  is set to 1. In addition, the corresponding Jacobians can be calculated exactly, so there is no need to apply SPSA to approximate them.

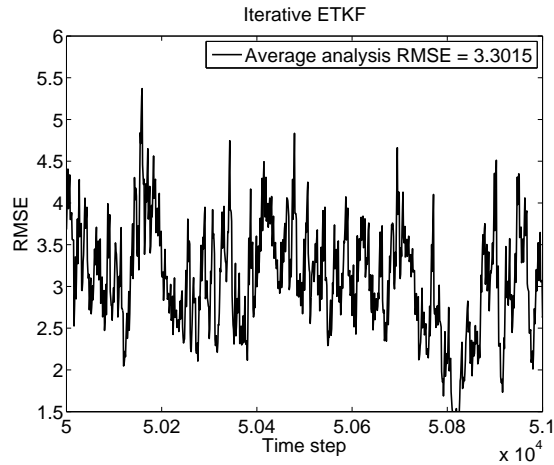




(a) Time series of residual norm

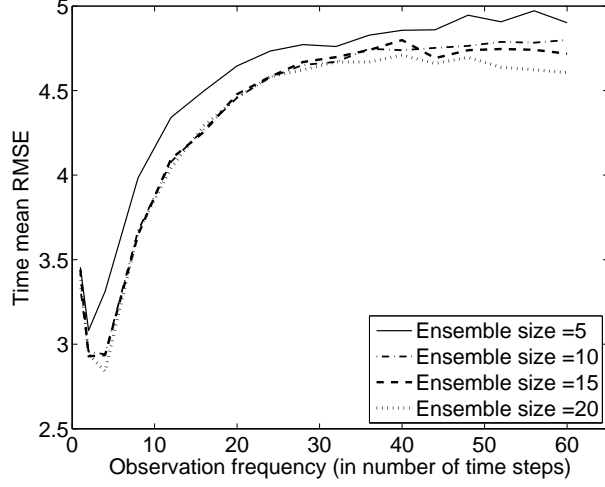


(b) Residual norm reduction of the iteration process at time step 50000

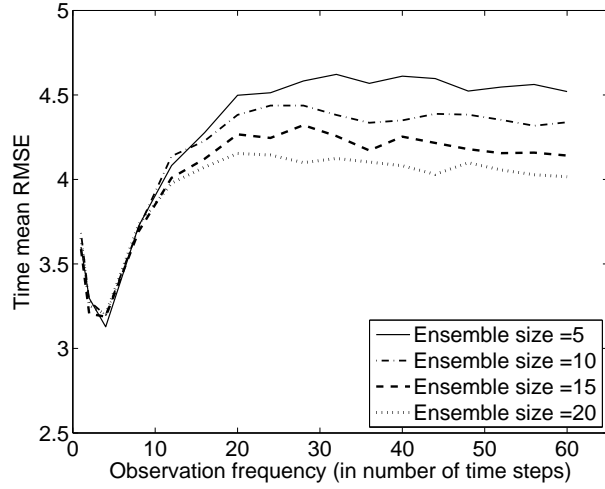


(c) Time series of RMSE

FIG. 7. As in Figure 4, except that the length of the assimilation time window is now 100000 time steps. For visualization, in the upper and lower panels we show only the time series between the time steps 50000 and 51000. The average time mean RMSE in the legend of the lower panel is, however, calculated with respect to the whole assimilation time window.



(a) Variances of observation errors are 0.01



(b) Variances of observation errors are 10

FIG. 8. Time mean RMSEs of the IETKF-RN as functions of the observation frequency (in number of time steps), for different ensemble sizes. Variances of observation errors in the upper panel are 0.01, and those in the lower panel are 10.

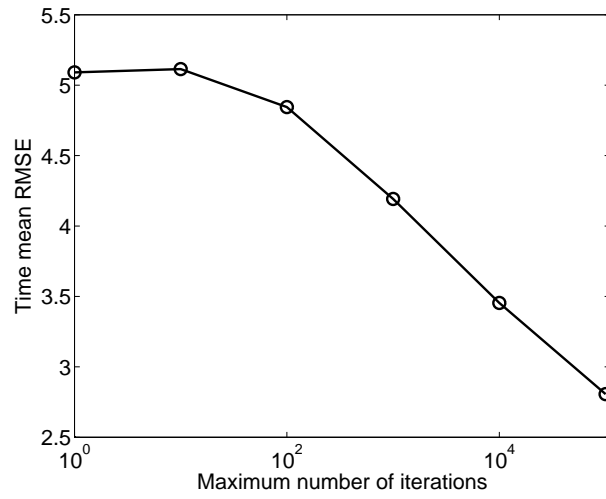


FIG. 9. Time mean RMSEs of the IETKF-RN as a function of the maximum number of iterations. With different orders of the magnitudes of the maximum iteration numbers, the logarithmic scale is used for the horizontal axis.

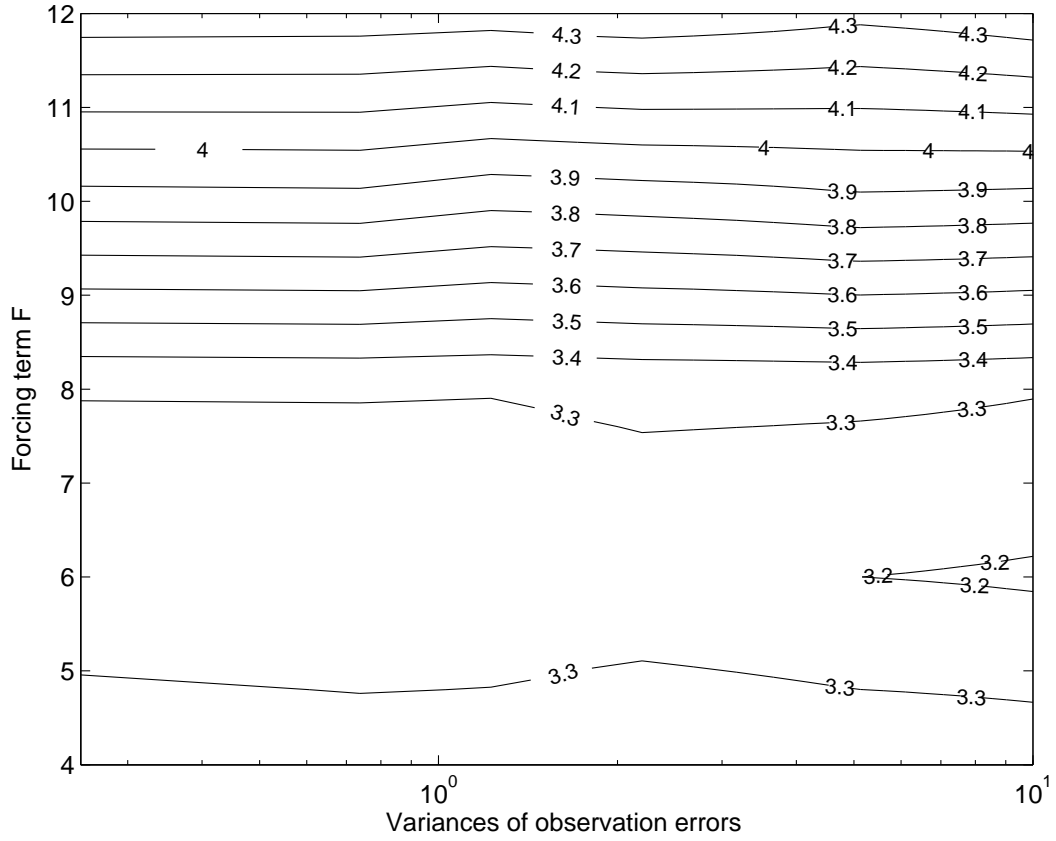
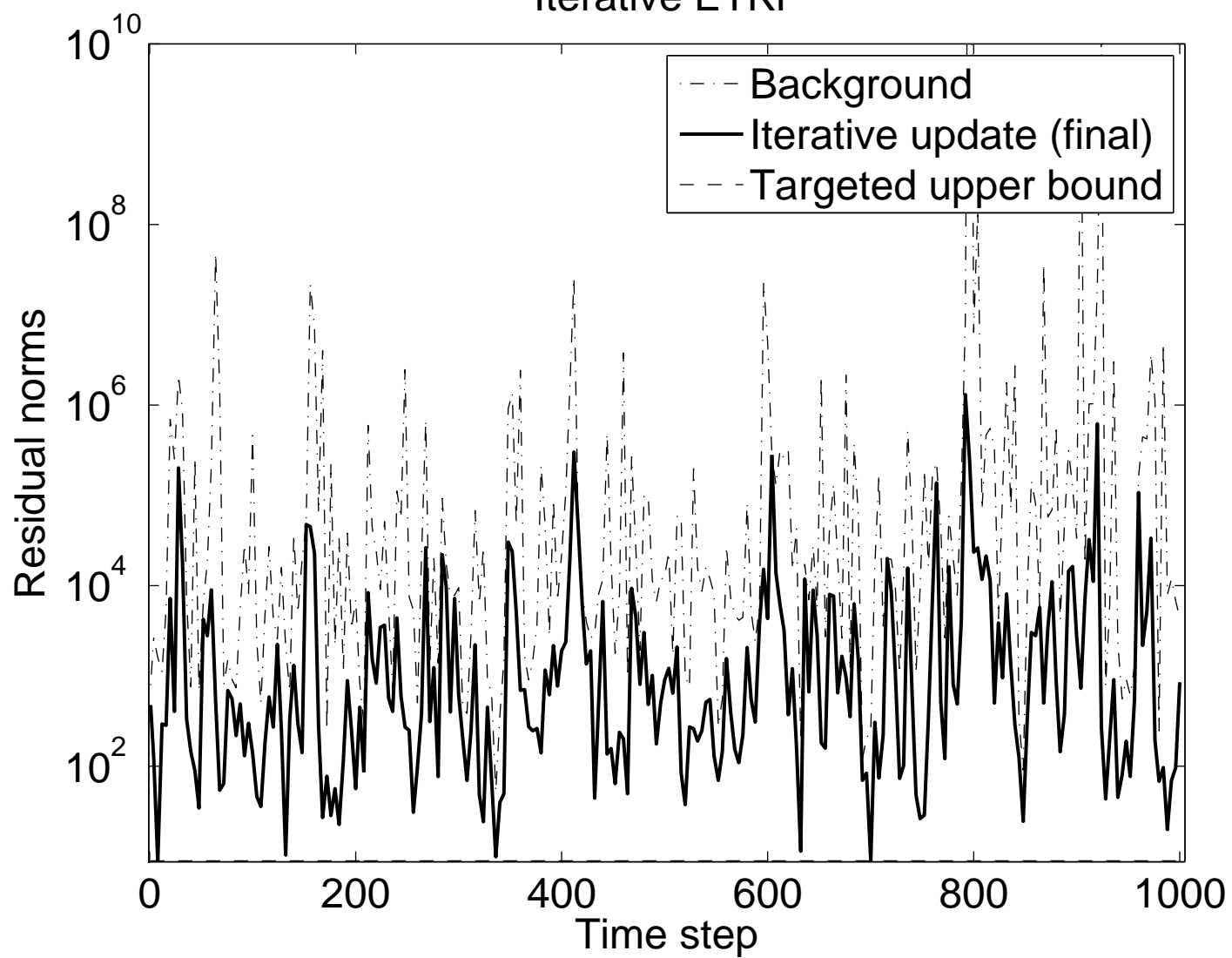
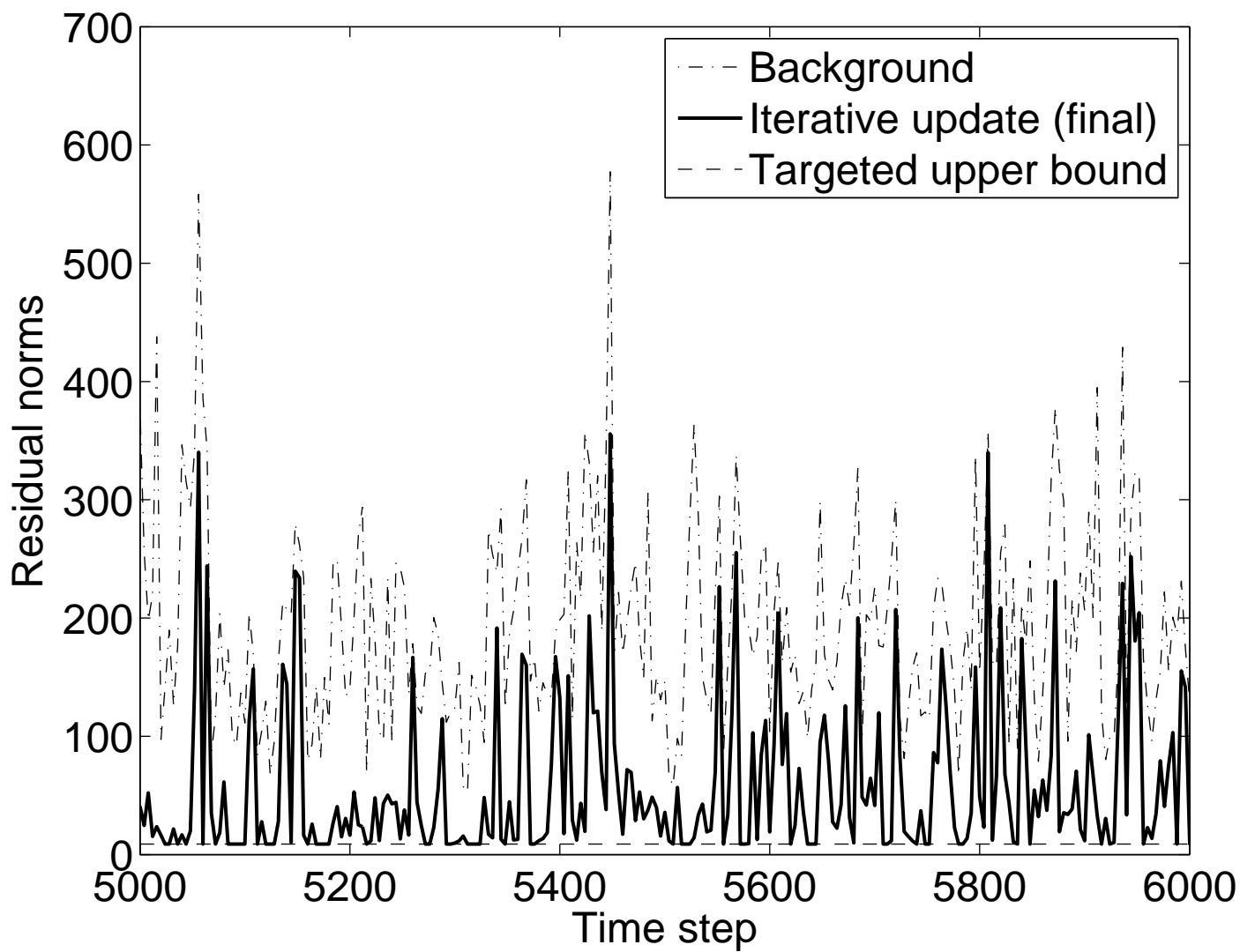


FIG. 10. Time mean RMSE of the IETKF-RN as functions of the potentially mis-specified forcing term  $F$  and the variances of observation errors. For visualization, here the logarithmic scale is also used for the horizontal axis.

# Iterative ETKF





# Iterative ETKF

